International Journal of Engineering Issues
Vol. 2015, no. 2, pp. 65-72
ISSN: 2458-651X

Rainfall Frequency Analysis Using L-moments of Probability Distributions

N. Vivekanandan
Central Water and Power Research Station, Pune, Maharashtra, India
E-mail: *anandaan@rediffmail.com*

Abstract **-** Frequency analysis of rainfall would enhance the management of water resources applications as well as the effective utilization of water resources. This can be carried out by fitting of probability distribution to the series of annual 1-day maximum rainfall data. In this paper, method of L-moments (LMO) is used for determination of parameters of Generalized Extreme Value (GEV), Gumbel (EV1) and Frechet (EV2) probability distributions. The adequacy of fitting of probability distribution adopted in rainfall frequency analysis is evaluated by applying Goodness-of-Fit (GoF) tests viz., Chi-square and Kolmogorov-Smirnov and diagnostic test (using D-index). Based on GoF and diagnostic test results, the study identifies the EV1 is better suited probability distribution for estimation of rainfall at Narwar whereas GEV for Banswara.

Keywords: *Chi-square; D-index; Generalized Extreme Value; Gumbel; Kolmogorov-Smirnov; L-moments; Rainfall*

I. INTRODUCTION

Rainfall is one of the most important input parameter to crop production and its occurrence and distribution is erratic, temporal and spatial variations in nature. Most of the hydrological events occurring as natural phenomena are observed only once. One of the important problem in hydrology deals with the interpreting past records of hydrological event in terms of future probabilities of occurrence [1]. Under these circumstances, Rainfall Frequency Analysis (RFA) would enable us to determine the expected rainfall for a given return period. This information can be used to prevent floods and droughts, and applied to planning and designing of water resources related to engineering such as reservoir design, flood control work and soil and water conservation planning [2].

Out of number of probability distributions, Extreme Value Distributions (EVDs) include Generalized Extreme Value (GEV), Gumbel (EV1) and Frechet (EV2) distributions are generally applied for RFA. EVDs arise as limiting distributions for the sample of independent, identically distributed random variables, as the sample size increases. Generally, Method of Moments (MoM) is used for determination of parameters of the distributions [3]. But, the MoM is not giving satisfactory results though the method exists for a longer period. It is sometimes difficult to assess exactly what information about the shape of a distribution is conveyed by its moments of third and higher order; the numerical values of sample moments particularly when the sample is small, can be very different from those of the probability distribution from which the sample was drawn; and the estimated parameters of distributions fitted by the MoM are often less accurate than those obtained by other estimation procedures such as maximum likelihood method, method of least squares and probability weighted moments. To overcome this, the alternative approach, namely L-moments (LMO) is discussed in this paper and also used in RFA [4].

In the recent past, number of studies has been carried out by different researchers on adoption of probability distributions for RFA. Topaloglu [5] reported that the frequency analysis of the largest, or the smallest, of a sequence of hydrologic events has long been an essential part of the design of hydraulic structures. Guevara [6] carried out hydrologic analysis using probabilistic approach to estimate engineering design parameters of storms in Venezuela. Kumar and Chatterjee [7] employed the LMO to define homogenous regions within 13 gauging sites of the north Brahmaputra region of India. Di Balldassarre et al. [8] used the LMO for regionalization of annual precipitation in

northern central Italy. Eslamian and Feizi [9] carried out EVA using monthly maximum rainfall for an arid region in Isfahan Province (Iran) through LMO. Gonzalez and Valdes [10] applied LMO for regionalization of monthly rainfall in the Jucar River basin. Yurekli et al. [11] found GEV and 3-parameter Log-Normal (LN3) distributions (using LMO) as the regional distribution functions for the maximum daily rainfall of Cekerec watershed, Turkey. Gubareva and Gartsman [12] analysed the extreme hydrometeorological characteristics adopting GEV, Generalised Pareto, LN3 and Pearson distributions through LMO. Badreldin and Feng [13] carried out the regional RFA for the Luanhe Basin, Hebei-China by using LMO and Cluster Techniques. But there is no general agreement in applying particular distribution for RFA for any region or country. Moreover, when different distributional models are used for modelling of rainfall data series, a common problem that arises is how to determine which model fits best for a given set of data. This can be answered by formal statistical procedures involving Goodness-of-Fit (GoF) and diagnostic tests; and the results are quantifiable and reliable. GoF tests such as Chi-square (χ^2) and Kolmogorov-Smirnov (KS) are applied for checking the adequacy of fitting of probability distributions to the recorded rainfall data. A diagnostic test of D-index is used for the selection of most suitable probability distribution for estimation of rainfall. The procedures adopted in RFA using probability distributions and computation of GoF and diagnostic tests are described in the following sections.

II. METHODOLOGY

LMO is analogous to the conventional moments but can be estimated linear combination of order statistics, i.e., by L-statistics. LMO is less subject to bias in estimation and approximate their asymptotic normal distribution more closely in finite samples.

A) Theoretical Description of LMO

Method of LMO is a modification of the probability weighted moments method explored by Hosking and Wallis [14]. Parameters of the distribution are estimated by equating the sample LMO (l_r) with the distribution of LMO (b_r). In practice, LMO must be estimated from a finite sample. Let $R_{1n} \le R_{2n \le \ldots \le} R_{nn}$ be the ordered sample of size n. The sample LMO is given by:

$$l_{r+1} = \sum_{k=0}^{r} \frac{(-1)^{r-k}(r+k)!}{(k!)^2(r-k)!} b_k \quad (1)$$

where, l_{r+1} is the r+1th sample moment and b_k is an unbiased estimator of β_k with

$$b_k = n^{-1} \sum_{i=k+1}^{n} \frac{(i-1)(i-2).....(i-k)}{(n-1)(n-2).....(n-k)} R_{in} \quad (2)$$

The first two sample LMOs are expressed by:

$$l_1 = b_0 \text{ and } l_2 = 2b_1 - b_0 \quad (3)$$

Table 1 gives the details of quantile function and parameters of EV1, EV2 and GEV distributions (using LMO).

Table1. *Quantile function and parameters of probability distributions*

S.No.	***Distribution***	***Quantile function (R_T)***	***Parameters of distribution (using LMO)***
1	EV1	$R_T = \xi - \alpha \ln(-\ln(1-(1/T)))$	$\xi = l_1 - 0.5772157\alpha$; $\alpha = l_2 / \log 2$
2	EV2	$R_T = \alpha e^{(-\ln(-\ln(1-(1/T))))/k}$	By using the logarithmic transformation of the recorded data, parameters of EV1 are initially obtained by LMO; and used to determine the parameters of EV2 from $\alpha = e^{\xi}$ and k=1/(scale parameter of EV1).
3	GEV	$R_T = \xi + \alpha(1-(-\ln(1-(1/T)))^k)/k$	$z = (2/(3+t_3)) - (\ln 2/\ln 3)$; $k = 7.8590z + 2.9554z^2$; $\alpha = l_2 k/(1-2^{-k})\Gamma(1+k)$; $\xi = l_1 + (\alpha(\Gamma(1+k)-1)/k)$

In Table 1, ξ, α, k are the location, scale and shape parameters respectively; μ (or $\overline{R}$), σ (or S_R) and C_S (or Ψ) are the average, standard deviation and coefficient of skewness of the recorded data; P is the probability of exceedance; ϕ^{-1} is the inverse of the standard normal distribution function and $\varphi^{-1} = Z_P = (P^{0.135} - (1-P)^{0.135})/0.1975$; R_T is the estimated rainfall by probability distributions corresponding to return period (T); sign(k) is ± 1 depending on the sign of k [15].

B) Goodness-of-Fit Tests

GoF tests such as χ^2 and KS are applied for checking the adequacy of fitting of probability distributions to the series of recorded rainfall data. Theoretical description of GoF tests are as follows:

χ^2 test:

$$\chi^2 = \sum_{j=1}^{NC} \frac{(O_j(R) - E_j(R))^2}{E_j(R)} \tag{4}$$

where, $O_j(R)$ is the observed frequency value of j^{th} class, $E_j(R)$ is the expected frequency value of j^{th} class and NC is the number of frequency classes [16]. The rejection region of χ^2 statistic at the desired significance level (η) is $\chi_C^2 \geq \chi^2_{1-\eta, NC-m-1}$. Here, m denotes the number of parameters of the distribution.

KS test:

$$KS = \operatorname*{Max}_{i=1}^{N} (F_e(R_i) - F_D(R_i)) \tag{5}$$

Here, $F_e(R_i) = i/(n+1)$ is the empirical CDF of R_i in which 'i' is the rank assigned to the sample values arranged in ascending order and $F_D(R_i)$ is the computed CDF of R_i [17].

Test criteria: If the computed values of GoF tests statistic given by the distribution are less than that of the theoretical values at the desired significance level (η), then the distribution is found to be acceptable for RFA.

C) Diagnostic Test

The selection of most suitable probability distribution for RFA is performed through D-index test (USWRC, 1981), which is defined as below:

$$\text{D-index} = \left(1/\overline{R}\right) \sum_{i=1}^{6} \left| R_i - R_i^* \right| \tag{6}$$

Here, $\overline{R}$ is the average value of the recorded data whereas R_i and R_i^* are the six highest recorded and corresponding estimated values by probability distribution. The distribution having the least D-index is considered as better suited distribution for estimation of rainfall [18].

III. APPLICATION

In this paper, a study was carried out to evaluate the probability distributions adopted for EVA of rainfall for Narwar and Banswara. The AMR series was extracted from the daily rainfall data recorded at Narwar for the period 1988 to 2004 and Banswara for the period 1969 to 2012 and used for EVA. For Narwar, the data for the missing years were replaced with the series maximum value of the recorded data based on AERB guidelines and the series with imputed data used for EVA. For Banswara, there is no missing data in the rainfall series and therefore no data was imputed. Table 2 gives the descriptive statistics of AMR recorded at Narwar and Banswara rain gauge stations.

Table 2. *Descriptive statistics of AMR*

Rain gauge station	*Statistical parameters (SD: Standard Deviation; CV: Coefficient of Variation)*				
	Mean (mm)	*SD (mm)*	*CV (%)*	*Skewness*	*Kurtosis*
Narwar	143.9	76.0	52.8	1.243	0.419
Banswara	186.3	97.1	52.1	0.572	-0.841

IV. RESULTS AND DISCUSSIONS

The procedures described above for estimating 1-day maximum rainfall have been implemented adopting computer codes and used in RFA. The program gives the (i) parameters of GEV, EV1 and EV2 distributions (using LMO); (ii) estimated rainfall for different return periods; and (iii) GoF tests statistic and D-index values.

A) Estimation of Rainfall by Probability Distributions

The parameters of EV1, EV2 and GEV distributions were determined by LMO and used for estimation of rainfall at Narwar and Banswara sites; and the results are presented in Tables 3 and 4.

Table 3. *Estimated 1-day maximum rainfall at Narwar*

Return period (year)	*Estimated 1-day maximum rainfall (mm)*		
	EV1	*EV2*	*GEV*
2	131.8	117.8	121.4
5	196.9	175.7	182.8
10	240.0	228.9	234.5
20	281.3	294.9	294.4
50	334.8	409.5	390.3
100	374.9	523.8	478.7
200	414.8	669.2	584.1
500	467.5	924.7	755.3
1000	507.4	1180.8	914.3

Table 4. *Estimated 1-day maximum rainfall at Banswara*

Return period (year)	*Estimated 1-day maximum rainfall (mm)*		
	EV1	*EV2*	*GEV*
2	169.5	149.2	169.1
5	259.7	227.9	259.3
10	319.4	301.8	319.4
20	376.7	395.0	377.3
50	450.8	559.7	452.8
100	506.3	726.7	509.7
200	561.7	942.7	566.6
500	634.7	1328.8	642.2
1000	689.9	1722.4	699.6

From Table 3 and 4, it may be noted that the estimated rainfall using EV2 distribution is consistently higher when compared to the corresponding values of EV1 and GEV distributions for return period 20-year and above.

B) Analysis Based on GoF Tests

The adequacy of fitting of EV1, EV2 and GEV distributions were evaluated by applying GoF tests to the recorded rainfall data. The GoF tests statistic values were computed by the probability distributions through Eqs. (4) and (5), and the results are presented in Table 5.

Table 5. *Computed and theoretical values of GoF tests using EV1, EV2 and GEV* distributions

GoF tests	*EV1*		*EV2*		*GEV*	
	Computed	*Theoretical*	*Computed*	*Theoretical*	*Computed*	*Theoretical*
Narwar						
χ^2	11.471	5.990	4.882	5.990	7.706	3.840
KS	0.177	0.318	0.130	0.318	0.119	0.318
Banswara						
χ^2	8.091	9.490	14.364	9.490	8.091	7.810
KS	0.133	0.205	0.182	0.205	0.132	0.205

By applying the test criteria to the GoF tests results, as given in Table 5, it may be noted that:

i) The χ^2 test supported the use of EV2 distribution for RFA for Narwar whereas EV1 for Banswara.
ii) The KS test results confirmed that the EV1, EV2 and GEV distributions are acceptable for RFA for Narwar and Banswara.

C) Analysis Based on Diagnostic Test

For the selection of most suitable distribution for estimation of rainfall, D-index values were computed by EV1, EV2 and GEV distributions through Eq. (6) and the results are presented in Table 6.

Table 6. *D-index values of EV1, EV2 and GEV distributions*

Rain gauge station	*EV1*	*EV2*	*GEV*
Narwar	1.492	1.523	1.463
Banswara	0.523	1.501	0.536

From the diagnostic test results, as given in Table 6, it may be observed that:

i) The D-index values of GEV for Narwar and EV1 for Banswara were found as minimum when compared to the corresponding values of other probability distributions.
ii) χ^2 test results didn't support use of GEV distribution for RFA for Narwar though the D-index value of the distribution is observed as minimum.
iii) There is no much variation between the D-index values computed from EV1 and GEV distributions for Banswara.

From the above, it may be noted that the quantitative assessment made through GoF and diagnostic tests are inconclusive. Therefore, the selection of most suitable distribution for estimation of rainfall was made through qualitative assessment using probability plots, as presented in Figures 1 and 2.

Figures 1 and 2, it can be seen that the fitted curves using EV1 (for Narwar) and GEV (for Banswara) distribution are in the form of linear and very close to the line of agreement with the recorded data. Based on qualitative and quantitative assessment, it was identified that the EV1 is better suited probability distribution for estimation of rainfall for Narwar whereas GEV for Banswara. By considering the design-life of the structure over the entire intended economic life time, the study recommended that the 1000-year return period estimated rainfall of about 507 mm for Narwar (using EV1) and 700 mm for Banswara (using GEV) could be used for design purposes.

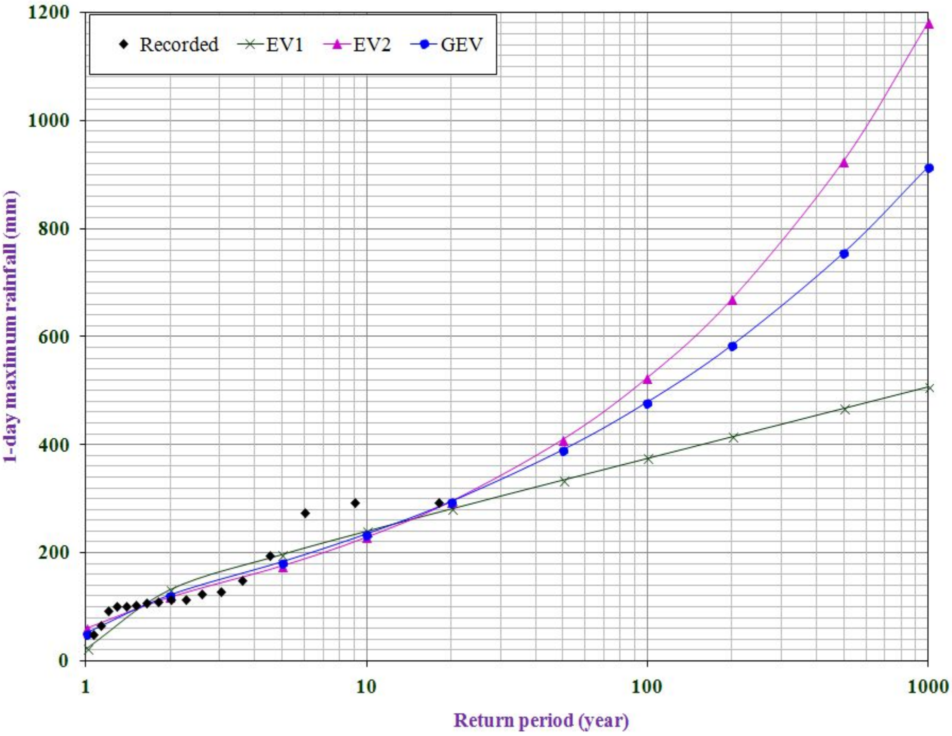

Figure 1. *Plots of recorded and estimated rainfall using EV1, EV2 and GEV distributions for Narwar*

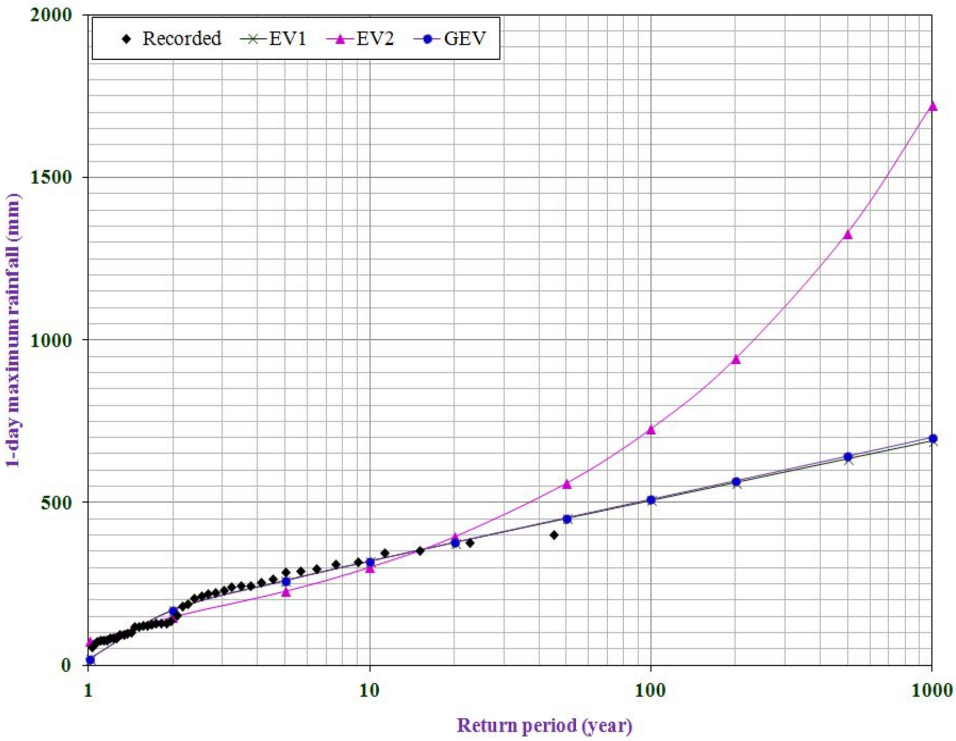

Figure 2. *Plots of recorded and estimated rainfall using EV1, EV2 and GEV distributions for Banswara*

V. CONCLUSIONS

The paper describes the study on evaluation of EV1, EV2 and GEV distributions for RFA for Narwar and Banswara through GoF and diagnostic tests. The following conclusions are drawn from the study:

i) For return period 20-year and above, it is observed that the estimated rainfall by EV2 (MLM) is consistently higher than the corresponding values of EV1 and GEV distributions for Narwar and Banswara.

ii) The χ^2 test results support the use of EV2 distribution for EVA of rainfall for Narwar whereas EV1 for Banswara.

iii) The KS test results confirm the EV1, EV2 and GEV distributions are found to be acceptable for RFA for Narwar and Banswara.

iv) Based on qualitative and quantitative assessment, the study identifies the EV1 distribution is better suited for estimation of rainfall at Narwar whereas GEV for Banswara.

v) The study suggests the 1000-year return period estimated rainfall of about 507 mm for Narwar (using EV1) and 700 mm for Banswara (using GEV) could be used as the design parameter for planning and design of hydraulic structures.

ACKNOWLEDGEMENTS

The author is grateful to the Director, Central Water and Power Research Station, Pune, for providing the research facilities to carry out the study. The author is thankful to M/s Nuclear Power Corporation of India Limited, Mumbai and India Meteorological Department, Pune, for supply of rainfall data.

REFERENCES

[1] BHAKAR, S. R., BANSAL, Anil Kumar, CHHAJED, Neeraj, *et al.* Frequency analysis of consecutive days maximum rainfall at Banswara, Rajasthan, India.*ARPN Journal of Engineering and Applied Sciences*, 2006, vol. 1, no 3, p. 64-67.

[2] DABRAL, P. P., PAL, Mautushi, et SINGH, R. P. Probability analysis for one day to seven consecutive days annual maximum rainfall for Doimukh (Itanagar), Arunachal Pradesh. *Journal of Indian Water Resources*, 2009, vol. 2, p. 9-15.

[3] RAO, A.R., et HAMEED, K.H. Flood frequency analysis, CRC Press, Boca Raton, Florida (USA), 2000.

[4] HOSKING, J.R.M. L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *Royal Statistical Society, Series-B*, 1990, Vol. 52, No. 1, p. 105-124.

[5] TOPALOGLU, F. Determining suitable probability distribution models for flow and precipitation series of the Seyhan River basin, *Turkish Journal of Agriculture and Forestry*, 2002, vol. 26, no. 1, p. 189 – 194.

[6] E. Guevara, Engineering design parameters of storms in Venezuela, Hydrology Days, pp. 80-91, 2003.

[7] KUMAR, Rakesh et CHATTERJEE, Chandranath. Regional flood frequency analysis using L-moments for North Brahmaputra region of India. *Journal of Hydrologic Engineering*, 2005, vol. 10, no 1, p. 1-7.

[8] BALDASSARRE, G. Di, CASTELLARIN, A., et BRATH, A. Relationships between statistics of rainfall extremes and mean annual precipitation: an application for design-storm estimation in northern central Italy. *Hydrology and Earth System Sciences*, 2006, vol. 10, no 4, p. 589-601.

[9] ESLAMIAN, S. Saeid et FEIZI, Hussein. Maximum monthly rainfall analysis using L-moments for an Arid Region in Isfahan Province, Iran. *Journal of Applied Meteorology and Climatology*, 2007, vol. 46, no 4, p. 494-503.

[10] GONZÁLEZ, Javier et VALDÉS, Juan B. A regional monthly precipitation simulation model based on an L-moment smoothed statistical regionalization approach. *Journal of Hydrology*, 2008, vol. 348, no 1, p. 27-39.

[11] YUREKLI, Kadri, MODARRES, Reza, OZTURK, Fazli, *et al.* Regional daily maximum rainfall estimation for Cekerek Watershed by L-moments.*Meteorological Applications*, 2009, vol. 16, no 4, p. 435-444.

[12] GUBAREVA, T. S. et GARTSMAN, B. I. Estimating distribution parameters of extreme hydrometeorological characteristics by L-moments method. *Water resources*, 2010, vol. 37, no 4, p. 437-445.

[13] BADRELDIN, G. H. H. et FENG, P. Regional rainfall frequency analysis for the Luanhe Basin using L-moments and cluster techniques. In : *International Conference on Environmental Science and Development (ICESD 2012)*. 2012. p. 126-135.

[14] HOSKING, J. R. M. et WALLIS, J. R. Some statistics useful in regional frequency analysis. *Water Resources Research*, 1993, vol. 29, no 2, p. 271-281.

[15] HOSKING, Jonathan Richard Morley et WALLIS, James R. *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press, 2005.

[16] ANNIS, C. Goodness-of-fit tests for statistical distributions. *Statistical Engineering (www. statisticalengineering. com)*, 2007.

[17] ZHANG, Jin. Powerful goodness-of-fit tests based on the likelihood ratio.*Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2002, vol. 64, no 2, p. 281-294.

[18] United States Water Resources Council (USWRC), Guidelines for determining flood flow frequency, Bulletin No. 17B, 1981.